On Paul Ford's "Moral AI Companies"

Rob Tow
Nova Lux, New Mexico, USA, Sol III
26 April 2026

Paul Ford's New York Times essay on "moral AI companies" has one durable point under its kaiju frosting: ethics do not scale once they leave the founder's head. The point is correct; Ford's metaphors are not. AI deployment is not a single commodity on a uniform grid but a nested ecology of coupled OODA loops, where the feedback delays Wiener would name and the symbiosis-and-parasitism Margulis and Kropotkin would diagnose are doing real work. Spam filtering is a better ancestor for governance than the electric grid, and layer-specific obligations are a better frame than utility regulation. The nearer danger is closer to Forbidden Planet *than to Godzilla.*

Paul Ford's *Times* essay this Sunday morning on "moral AI companies"(gift link) is useful, but only after one scrapes off the Frosting. Its real argument is not the Godzilla conceit, the acronym play, or the Sunday-opinion charm, but rather that ethics do not scale when left inside the corporate boundary. That is correct, and worth saying. But the essay keeps reaching for metaphors that are too cute, too blunt, or too mechanically simple for the machine under discussion. It makes maps that are too cartoonish to navigate the actual territory, altho'they are charmingly colorful.

The kaiju conceit is precious. It is also irrelevant to sober technological planning. AI firms are not Godzilla with a mission statement. They are vertically tangled capability providers: data-center operators, model builders, API vendors, search mediators, software generators, publishing engines, labor-market accelerants, surveillance-adjacent inference systems, and, in some cases, state-adjacent contractors. The problem is not whether the beast is friendly. The problem is which control loops exist, who can inspect them, who

can interrupt them, which incentives they optimize, and who bears the cost when a model-mediated output becomes institutional reality. What is designed? What is emergent? Who is responsible for either? These are dull questions only to those who prefer upholstery to architecture.

His electric-utility analogy is exceedingly weak. Electric utilities deliver a comparatively uniform commodity through a physically bounded grid. You can meter kilowatt-hours, inspect substations, regulate interconnection, measure reliability, and investigate a blackout as a physical-system failure. AI companies do not provide one commodity. They provide shifting bundles of capability: text generation, code synthesis, search substitution, tutoring, fraud automation, legal drafting, medical summarization, image generation, targeting support, bureaucratic acceleration, and emotional simulation. The same model can be harmless as a toy and dangerous inside a claims-denial pipeline, a hiring filter, a drone-support workflow, or a hospital documentation system. This is not a grid. It is a set of capability stacks embedded in overlapping ecologies.

That ecological character matters. At the bottom are data centers, power contracts, water demand, chips, networking, cooling, and cloud orchestration. Above that are training corpora, tokenization, pretraining, post-training, reinforcement learning or preference tuning, safety filters, retrieval layers, tool use, agents, API policies, application wrappers, customer workflows, logging, monitoring, and incident response. Around all of it are users, attackers, publishers, governments, advertisers, competitors, litigants, open-source projects, military customers, institutional buyers, and ordinary people trying to get through the day. These actors do not stand politely in a diagram. They observe, adapt, exploit, cooperate, defect, and reorient, while harms enter at one layer and propagate through the others. The system evolves, rapidly, and without asking permission

from the metaphor department.

This is where John Boyd is more useful than Godzilla. Boyd, the fighter pilot and military theorist, is usually compressed into the phrase "OODA loop": observe, orient, decide, act. That compression is useful but incomplete. Boyd's deeper point was that conflict is not a duel between static machines, but a contest among adaptive systems trying to perceive, interpret, act, and re-perceive faster and more coherently than their opponents. AI deployment is now exactly such an ecology of coupled OODA loops, in which every visible rule becomes a terrain feature and every terrain feature invites maneuver. Spammers learn the classifier, publishers deform prose to satisfy search and recommendation systems, students and institutions adapt around detection, companies reorganize workflows around automation, and fraudsters patiently discover what the model will say, what the filter will miss, and what the user can be induced to believe. Regulators, meanwhile, tend to observe late, orient through stale categories, and act after the adversarial population has already mutated. The result is not a product-risk curve in the ordinary corporate sense, but an evolving field of coupled and nested decision loops.

This is also why Norbert Wiener belongs in the discussion. AI governance is a cybernetic problem before it is a branding problem. It concerns feedback, delay, amplification, classification, control, and the pathologies that appear when a system's outputs become part of its future inputs. A model affects the world, the changed world produces new data, the new data trains or tunes the next model, and institutions begin to adapt around the machine's affordances. Once that loop closes, "the model"is no longer a discrete object. It is a participant in a control system, which is to say, a machine in the old and serious sense: not merely gears and housings, but an organized relation among causes, signals, delays, feedbacks, and consequences.

The participants are nested at different scales. Some behave like mitochondria inside larger cells, some like eukaryotic cells, some like vertebrates browsing through forests, and some like forest ecosystems themselves. That is a major point that I see completely missing in all public discourse. The unit of analysis keeps shifting, and any policy that pretends otherwise is not policy but administrative stage scenery in a play to amuse and excite the masses.

Nor is the system merely competitive. Here Lynn Margulis and Peter Kropotkin sharpen the frame. Evolution is not only red-clawed competition; it is also symbiosis, mutualism, parasitism, niche construction, and cooperation under constraint. Publishers cooperate with search while gaming it. Platforms protect users from spam while harvesting their behavior. Open-source communities strengthen the field while also diffusing capability. Enterprise customers seek productivity while quietly moving institutional judgment into opaque machine layers. Model providers license serious sources while feeding on the ambient commons. The ecology is not good or bad in a simple moral sense. It evolves, and it selects. It rewards some behaviors, suppresses others, and changes the environment in which the next generation of behavior appears. New entities emerge from cooption, cooperation, and competition. Some are red in tooth and claw, and some are very charming. Nature has been making such jokes longer than Silicon Valley has been printing hoodies.

That is why the history of the spam-filtering lineage in email is a much better ancestor than the electric-grid analogy. It is fashionable now to sneer at "AI slop" and "stochastic parrots," often for good reason, but that fashion can obscure a harder engineering fact: machine learning and adjacent statistical methods have been quietly working below the hood at industrial scale for decades. Email spam filtering, including Bayesian filtering and machine learning, is the clearest example. It is not a fantasy of superintelligence. It is not a TED-talk

prophecy. It is machinery, deployed in commerce, against adaptive adversaries, at planetary scale.

This is not merely analogy by taste. I have run my own email server on a Linux hosting service for over twenty-five years. I have lived with this machinery under the hood: SpamAssassin scores, Bayesian training, SPF, DKIM, DMARC, blocklists, allowlists, forged headers, poisoned reputations, gray areas between bulk mail and fraud, and the constant low-grade evolutionary pressure of people trying to get garbage through the filter. This is not abstract to me. It is a working engineering ecology, built from ugly compromises, open-source tools, statistical inference, authentication protocols, local tuning, and scar tissue. It works well enough that most people never see the miracle. They only notice when the filter fails, at which point they discover that civilization apparently depends on a quite unromantic stack of rules, probabilities, signatures, and grudges.

I have been a cold-eyed rocket engineer at this level. The invisibility of this low-level engineering matters. Users do not experience the classifier; they experience the absence of garbage. When spam filtering works, it becomes part of the epistemic plumbing. When it fails, email becomes unusable. This is a better ancestor for present AI governance than electric utilities because it already contains the crucial structure: sender adaptation, classifier response, reputation systems, authentication protocols, user feedback, false positives, false negatives, economic pressure, distribution shift, and constant mutation by opponents. The enemy is not load demand. The enemy is other adaptive intelligences trying to live inside the classifier's blind spots. It is actually very biological.

This lineage runs through search ranking, recommender systems, fraud detection, ad targeting, malware classification, abuse moderation, and now generative AI. These are not merely technical services. They are nested selection environments, matryoshka dolls of incen-

tives and adaptations. Each creates incentives, which produce behavior, which becomes data, which retunes the system, which changes the incentives again. Competition and cooperation are both present, often in the same actor. A publisher cooperates with search while gaming it. A platform protects users from spam while harvesting their behavior. A model provider licenses serious sources while feeding on the ambient commons. An enterprise customer seeks productivity while quietly transferring institutional judgment into an opaque machine layer.

Classification, in such a system, is governance. A provider deciding that a message is spam is not merely performing hygiene. It is silently managing the boundary between signal and noise, commerce and fraud, speech and abuse, attention and manipulation. When this works, it disappears from ordinary view. Users do not experience the classifier; they experience the absence of garbage. That invisibility is politically important, because invisible infrastructure is where power hides, and because our species has a touching habit of calling invisible power "neutral" when it flatters us. This is where power resides, and where one hears Captain Renault in *Casablanca*: "I'm shocked, shocked to find that gambling is going on in here!"

Ford is on firmer ground when he says ethics do not scale. That is the real essay, buried under the monster suit, for which I can find some respect. Founder ethics are not governance. A small group can begin with sincere moral ambition, but once capital, lawsuits, procurement, state interest, compliance departments, platform scale, and competitive panic enter the loop, the original ethical intention becomes one input among many. Mission becomes rhetoric. Metrics become command. "Doing good" becomes a membrane around power. Then evolution occurs, which is seldom courteous enough to preserve the founding brochure; Google's "Don't be evil" survived mostly as a fossilized moral slogan.

The effective altruism material is partly relevant, but not as a primary technical explanation. Sam Bankman-Fried is useful mainly as a cautionary specimen: moral leverage without epistemic humility. Once a person believes that hypothetical future benefit is large enough, ordinary constraints begin to look like friction imposed by lesser minds. That is not engineering judgment. It is speculative moral accounting with other people's money, other people's institutions, and other people's future as collateral. One need not be a theologian to recognize indulgences when they are sold in mathematical dress. The same failure mode can appear in AI when distant catastrophe is used to justify present extraction, secrecy, copyright disregard, labor disruption, or regulatory evasion. The issue is not that long-term risks are imaginary. Some are very real. The issue is that a man who appoints himself steward of the far future can become remarkably casual about the near present, especially when the sacrifices are made by other people.

The celebrity roll call is also a distraction. Sam Altman, Elon Musk, Sam Bankman-Fried, and the rest are not irrelevant, but they are not the engineering substrate. They are executives, financiers, ideologues, brand animals, and competitive actors embedded in incentive fields. An engineer does not ask whether such men have good intentions. He asks what authority they possess, what feedback they receive, what they can externalize, what they are rewarded for ignoring, what constraints bind them when they are wrong, and whether failure kills the pilot, the shareholders, the bystanders, or merely the truth. The moral weather inside a founder's head is a poor substitute for a pressure gauge.

"Google zero"is another weak spot. The complaint smuggles in an entitlement: because Google once routed traffic to websites, those websites are somehow owed traffic forever. But a large fraction of the ad-supported web is not a noble commons. It is SEO chum,

affiliate-link farms, recipe sites padded with autobiographical sawdust, engagement traps, AI-generated filler, and advertising machinery wrapped around thin information. I do not regard the collapse of that ecosystem as a tragedy. Some wetlands are wetlands. Some are mosquito ditches beside a tire fire. We need not preserve every bog merely because something profitable breeds there.

The better distinction is between traffic-dependent slop and genuine knowledge-producing institutions. A recipe spam farm dying is not a civilizational wound. A local newspaper dying is. A specialist forum going dark is. So is a technical blog written by someone who actually knows the machinery. An independent historian, cook, mechanic, legal explainer, scientist, or working engineer losing discoverability is a real loss. The public problem is not that every page view must be preserved. The public problem is whether AI firms destroy the economic basis for producing reliable, original, inspectable human knowledge while feeding future systems on the decaying remains of attention bait.

This is where a liberal-arts analogy becomes central, not decorative. A serious liberal arts education was never a random scrape of all available text. It was curated. One encountered durable works, primary sources, hostile arguments, mathematics, history, languages, teachers, traditions, and enough canonical pressure to prevent mistaking the chatter of the day for the structure of the world. Plato, Aristotle, Aquinas, Machiavelli, Hobbes, Locke, Hume, Darwin, Marx, Nietzsche, Freud, Wiener, Arendt, and the rest were not encountered as paste. They were encountered as a sequence of argument, inheritance, revolt, correction, and misreading. The order mattered; the opposition mattered; the voice of the teacher mattered. And the marginal note mattered.

That is also the answer to the lazier version of "model collapse." Model collapse should not be treated as an argument for preserving

the whole web as an indiscriminate feedlot. In the technical sense, it is a distributional and epistemic problem: systems trained increasingly on synthetic, derivative, homogenized, or self-referential outputs can lose diversity, amplify errors, narrow their representation of the world, and become more confident about less. But the answer is not to preserve every ad-choked content farm so the model has something to chew. That is mistaking calories for nutrition, which is also how one gets both gout and bad prose.

A library is not a landfill, and a curriculum is not a scrape. A serious corpus should have provenance, genre, date, authorship, rank, contradiction, domain boundaries, and memory of why some works matter more than others. It should distinguish primary documents from commentary, peer-reviewed literature from press release, satire from assertion, law from blog post, propaganda from analysis, fiction from witness, and technical documentation from tutorial sludge. If AI companies want to build systems that mediate knowledge, they cannot treat the web as a quarry and then complain when the quarry fills with tailings. They should ask how universities, libraries, seminar traditions, and editorial disciplines have preserved knowledge across generations. This would require humility, of course, which is not yet a commonly supported cloud feature.

The engineering problem, considered in the context of generational governance, is not "more data."It is source quality, provenance, evaluation, deployment control, and feedback discipline. What went into the model? Under what license? With what weighting? How was it filtered? What was removed? What was overrepresented? What was labeled? What downstream uses were tested? What red-team results were ignored? What incidents are reported? What logs are retained? What can users contest? What systems are allowed to call tools, spend money, write code, send messages, touch infrastructure, or influence legally significant decisions? What scaffolding do

we want around the thoughts and emotions of our children?

That is where regulation should bite. Not at the level of monster stories, and not through a simplistic single utility-style framework, but through layer-specific obligations: corpus provenance, environmental accounting, model-risk management, deployment-specific liability, audit rights, incident reporting, security standards, adversarial testing, procurement rules, professional exclusion for reckless actors, and penalties that reach executives rather than merely taxing shareholders after the fact. An examination of corporations as legal fictive people is relevant here, because one cannot sensibly discuss responsibility while pretending that limited liability, corporate personhood, and managerial insulation are mere background conditions rather than active design features. A fictive person is a useful legal contrivance, but it should not be allowed the conscience of a ghost and the appetite of a terminator.

Ford is right to distrust self-sanctifying companies. He is right that voluntary ethics fail under scale. But the sober planning question is not whether the monster is friendly, whether the grid analogy comforts Congress, or whether every recipe site is entitled to tribute traffic. The question is what adaptive loops the system creates, what behavior those loops select for, what failures become cheap, what responsibility becomes deniable, and whether any public institution can still observe, orient, decide, and act faster than the private systems now reshaping the terrain.

Rather than Godzilla, I see the nearer danger as closer to *Forbidden Planet*: we build vast data centers drawing power at planetary scale, couple them to systems we only partly understand, and then discover that we have given machinery to the monsters from the id. Like Dr. Morbius, we may end up crying, "Why haven't I seen this all along?"

Originally published on Facebook, 26 April 2026.